

Preserving Data Journalism: A Systematic Literature Review

Bahareh Heravi, Kathryn Cassidy, Edie Davis & Natalie Harrower

To cite this article: Bahareh Heravi, Kathryn Cassidy, Edie Davis & Natalie Harrower (2021): Preserving Data Journalism: A Systematic Literature Review, Journalism Practice, DOI: [10.1080/17512786.2021.1903972](https://doi.org/10.1080/17512786.2021.1903972)

To link to this article: <https://doi.org/10.1080/17512786.2021.1903972>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 31 Mar 2021.



[Submit your article to this journal](#)



Article views: 118







[View related articles](#)



[View Crossmark data](#)

Preserving Data Journalism: A Systematic Literature Review

Bahareh Heravi ^a, Kathryn Cassidy ^b, Edie Davis ^c and Natalie Harrower ^d

^aSchool of Information & Communication Studies, University College Dublin, Dublin, Ireland; ^bDigital Repository of Ireland, Trinity College Dublin, Dublin, Ireland; ^cThe Library, Trinity College Dublin, Dublin, Ireland; ^dDigital Repository of Ireland, Royal Irish Academy, Dublin, Ireland

ABSTRACT

News organisations have longstanding practices for archiving and preserving their content. The emerging practice of data journalism has led to the creation of complex new outputs, including dynamic data visualisations that rely on distributed digital infrastructures. Traditional news archiving does not yet have systems in place for preserving these outputs, which means that we risk losing this crucial part of reporting and news history. Following a systematic approach to studying the literature in this area, this paper provides a set of recommendations to address lacunae in the literature. This paper contributes to the field by (1) providing a systematic study of the literature in the fields, (2) providing a set of recommendations for the adoption of long-term preservation of dynamic data visualisations as part of the news publication workflow, and (3) identifying concrete actions that data journalists can take immediately to ensure that these visualisations are not lost.

KEYWORDS


data journalism; data-driven journalism; data visualisation; data visualization; digital preservation; digital archiving; software preservation

Introduction

“Journalism is the first rough draft of history” (widely attributed to Philip Graham, 1963), and the archives of news organisations are an indispensable source for research into global history. Traditional journalistic outputs are usually published in text and audiovisual format, with news organisations having a longstanding history of archiving and preserving these outputs on various media, for example, paper, tape, or hard disc drives, depending on the historical time period and the original format of the output. Similarly, memory institutions such as national libraries and archives generally hold large and long-standing newspaper archives.

In the past decade journalism has become more “quantitatively oriented” (Coddington 2015, 332), and increasingly incorporates “data journalism” – a practice which uses datasets, computational tools and algorithms to create news stories (Heravi and Lorenz 2020). The output of this type of journalism includes traditional text and audiovisual formats, but also includes data visualisations and/or news applications. These visualisations communicate key aspects of the story, and without them, the story is either incomplete or entirely

CONTACT Bahareh Heravi  Bahareh.Heravi@ucd.ie

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/17512786.2021.1903972>.

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

missing. The issue for sustained access is that the visualisations rely on digital infrastructures (code, apps, platforms) that are not being systematically preserved and sustained. If the code behind the visualisation breaks, then the visualisation disappears or renders an error. Research into the winners of the Data Journalism Awards shows that the majority of award-winning data journalism includes dynamic and interactive graphics, built on and hosted by a range of internal and external services (Ojo and Heravi 2018). Traditional news archiving has not accounted for these dynamic and interactive narratives. At the same time, an increasing number of such new, complex outputs are being generated in newsrooms across the world every day, and it is expected that this trend will continue to grow. Without intervention, we will lose a crucial part of reporting and news history.

Data journalism – and its enthusiastic uptake – has opened up a new set of challenges for preservation and created a requirement for new guidelines and practices. In this paper, we take a systematic approach to studying the literature in order to identify these challenges and propose potential solutions to preservation of data journalism outputs.

Problem Statement

Data visualisations are one of the core outputs of data journalism. They could be in the form of static image files (e.g., jpeg, gif, png, etc.), but in many cases, they are dynamically generated at the time of viewing, by computer code (e.g., JavaScript), and hosted online, often on platforms managed and maintained by third parties. These visualisations are often interactive, and anything beyond a simple image is considered a dynamic data visualisation. Such dynamic content cannot be captured by existing tools and methods of archiving, for example, tools for archiving web pages, and consequently are being lost (Boss and Broussard 2017; Broussard and Boss 2018; Hansen and Paul 2017). Dynamic data visualisations and news apps are essentially software, and their preservation, therefore, should follow methods for software preservation. As Broussard points out “software preservation is a specialised, idiosyncratic pursuit that requires more money and more specialized labor than is available at media organizations today” (2015b). Simply put, the organisation that creates the visualisation, and holds an interest in its preservation (the news organisation), is not usually the same organisation that holds the key to that visualisation’s sustainable accessibility.

In the language of digital preservation, dynamic data visualisations are considered “complex digital objects”. These differ from “simple” or “flat” objects such as image and video files, in that they are more challenging to maintain and preserve because they rely on complex digital infrastructures that contain a series of technical (inter)dependencies, where each part of the infrastructure must function in order to deliver the final output. Some of these infrastructures are maintained in-house by news organisations, whereas others rely on third party services and/or proprietary applications.

Simple objects are more likely to be maintained long term, because they fall under existing preservation methods used within news organisations since the beginning of the twentieth century. In contrast, the many infrastructures that support ongoing access to dynamic visualisations are not being systematically sustained or preserved in a way that would ensure access to data journalism outputs.

Evolving technology threatens preservation, but in a data journalism context, this is complicated by the range of systems involved. Below, we identify four primary factors that endanger the preservation of data journalism outputs, based on groupings of technologies:

Third-party Services

Many data visualisations make use of third-party data visualisation tools, such as Datawrapper and Flourish, which provide useful and often sophisticated assistance in creating visualisations.¹ However, the use of these tools creates risk because of dependencies on the tool provider: the tool may not be maintained by the provider, changes made to their underlying technologies may “break” the connection to published visualisation on a news site, or the service might disappear altogether. This has already come to pass with the shutdown of Silk.co (The Silk team 2017) and Google Fusion Tables (Melendez 2018), both data visualisation services once popular with data journalists. In the case of Silk.co, the website closed on short notice, ceasing access to any data visualisations that had not been exported or migrated by creators prior to the shutdown. A similar scenario happened a year later in December 2018 when Google announced that they would retire their Fusion Tables service. Fusion Tables were one of the tools behind many early examples of Data Journalism, such as the Wikileaks’ Iraq war logs published by the Guardian. The service was switched off at the end of 2019, and much of the associated content disappeared. The Guardian example depicted in Figure 1 is only one of many stories with missing visualisations across news organisations in the past number of years.

In-house Tools

While many workflows rely on third-party apps, some organisations have also designed in-house tools. These may afford greater control over the tool and its integration with internal technologies, but often these tools have been designed for specific purposes, such as to communicate the data behind a given data-driven piece. The longer-term

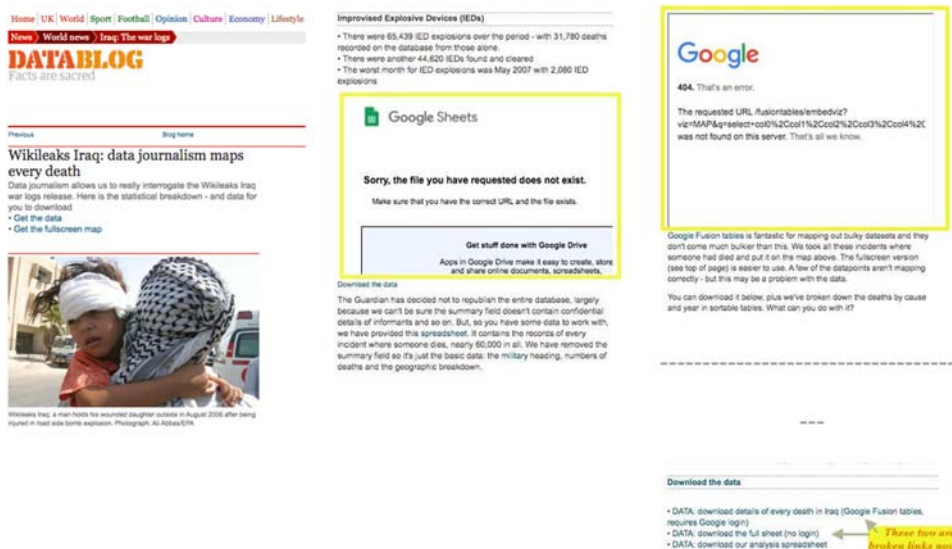


Figure 1. Screenshots from the Guardian story, depicting how the content gets lost when the third-party services are not maintained: www.theguardian.com/news/datablog/2010/oct/23/wikileaks-iraq-data-journalism. Screenshot taken on 5th August 2020.

use of the tool or its maintenance may not have been considered during the design process, or no strategy has been put in place to track, archive and preserve the output of such tools.

Content Management Systems

The public-facing website of a news organisation is usually fed by a back-end Content Management System (CMS), which itself needs to be maintained, updated, and periodically replaced by new platforms. Through these changes, the embedding functionality that connects the visualisation to the CMS can be broken or rendered incompatible. In this case, the visualisation and/or the tool remain intact, but the visualisation is not fetched or displayed properly on the news organisation website.

Myriad Other Technologies

While the above risks point to significant changes in known aspects of the technology chain, there are other dependencies that underpin visualisations, such as particular programming languages, libraries, databases, hosting platforms and tools. These change over time – by the news organisation, the tool provider, or globally – and changes can cause the data visualisation itself to no longer be accessible or viewable. An example of technological change can be seen in the consequences of Adobe's decision to deprecate Flash. In countless stories published around and before 2010, such as the Guardian's articles on Earthquakes, or the Financial Times' Banks' Earnings, the visualisation itself was the article. So their disappearance due to the deprecation of Flash resulted in empty pages, with the now-useless suggestions to download or update Flash Player (see [Figure 2](#)).

To highlight the extent of this matter even more, in 2010 Segel and Heer (2010) studied 58 visual stories from several publishing houses. At that time Flash was the go-to technology for creating interactive visualisations. Just 10 years past this study, Flash Player was deprecated and consequently very few of the visualisations remain accessible. Flash will not be the only casualty, as preferred apps and scripts continue to change over time.

In addition to the large-scale fails, all digital objects, simple or complex, are in danger of degradation or loss over time, due to factors such as data corruption (bit rot) – the obsolescence of file formats, software and hardware – and the limited lifespan of storage media. For all of these reasons, it is imperative that news media prioritise digital preservation.

This paper is motivated by the of lack of systematic research on best practices in the preservation of data journalism-related complex objects. The scope of this systematic literature review includes works relating to the preservation of dynamic data visualisation and associated software code and dynamic digital objects. The preservation of the datasets that underlie data visualisations is also key; in some cases, they are required to make the visualisation function as it is rendered, and in any case, the data should be persistently accessible in order to verify the findings communicated by the visualisation. However, this is a separate, larger issue for digital preservation (e.g., Berman (2008) and Lynch (2008)) and preservation of research datasets is being addressed elsewhere by international initiatives, such as the Research Data Alliance,² the CODATA committee of the International Science Council (Lin et al. 2020), the OECD (OECD 2007, 2017; Pilat and Fukasaku 2007), UNESCO (2020), and the European Commission particularly via the requirements

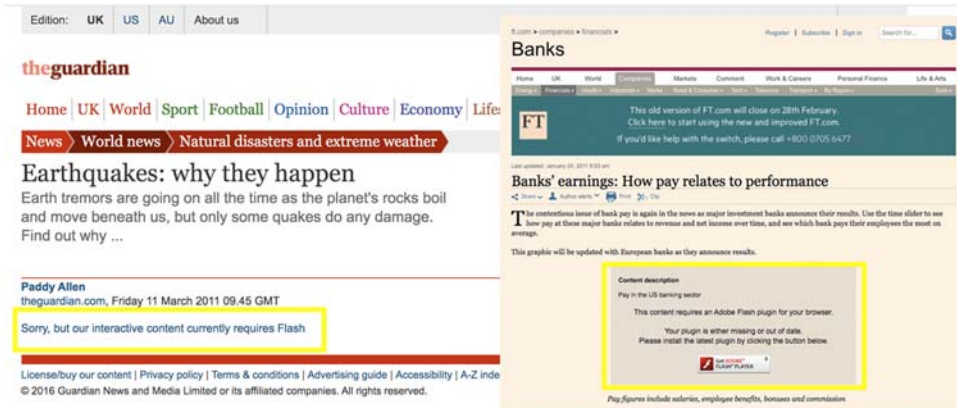


Figure 2. Screenshot from a Guardian story, depicting the disappearance of the full story due to the deprecation of Flash: <https://www.theguardian.com/world/interactive/2008/jan/23/earthquakes> and ig-legacy.ft.com/content/4ce7a094-1c9e-11df-8456-00144feab49a#axzz6murkxlLF. Screenshots taken on 21st February 2021.

for the European Open Science Cloud (Schoupe and Burgelman 2018; The European Commission Expert Group on FAIR Data 2018).

The primary aim of this paper is to study current and past practices of preserving both data journalism outputs and related digital objects, to assess what recommendations have been made for preserving such objects, and accordingly to provide a set of suitable recommendations for preservation of dynamic visual objects in data journalism.

Method

A systematic literature review is “a systematic, explicit, [comprehensive,] and reproducible method for identifying, evaluating, and synthesising the existing body of completed and recorded work produced by researchers, scholars, and practitioners” (Fink 2005, 3). Okoli (2015) notes that “a rigorous standalone literature review must be systematic in following a methodological approach, explicit in explaining the procedures by which it was conducted, comprehensive in its scope of including all relevant material, and, hence, reproducible by others who would follow the same approach in reviewing the topic” (880). The review protocol for this study was developed based on the recommendations from Kitchenham and Charters (2007) and Okoli (2015). Our adapted method is depicted in Figure 3.

Our approach in implementing this procedure is presented in detail in Appendix 1. This study focuses – by definition – on peer-reviewed publications. However, the professional communities have generated significant contributions relevant to this study. Therefore, to complement our systematic literature review, we further draw on a set of key non-academic contributions to the field in the Discussion and Recommendations sections.

Findings

This section presents the findings through studying the challenges associated with sustained access to dynamic data visualisations, the methods of preservation implemented

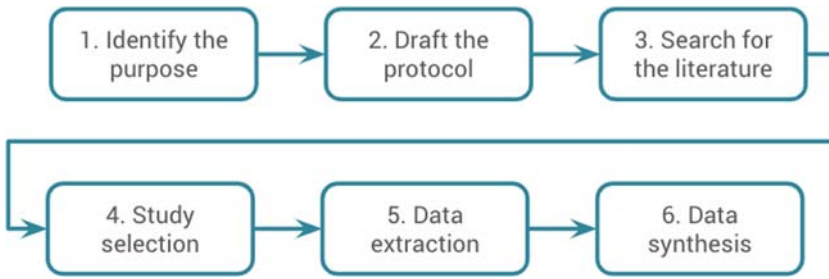


Figure 3. Literature review procedure, adapted from Kitchenham and Charters (2007) and (Okoli 2015).

or recommended in the literature, and the typical recommendations made in this field to combat associated challenges and relevant problems.

Challenges Associated with Sustaining Access to Dynamic Data Visualisations

The main themes that emerged in this study are illustrated in Figure 4, with the most commonly discussed challenges relating to *obsolescence* and *dependency*. These include file format obsolescence (where a particular format is no longer readable because it is not actively supported by contemporary software/hardware), changes to technical infrastructure, problems such as “bit rot”, which is the slow decay or corruption of digital data due to the accumulation of errors on a storage device (Monnens et al. 2009), or “dependency hell”, where multiple shared software libraries required by different parts of the code, in turn, rely on different versions of another library, creating conflicting dependencies on different versions (Harris 2013). In effect, these are not challenges to preservation *per se*, but rather they are central drivers of preservation actions to prevent the disappearance or breakage of dynamic data visualisations over time.

This set of challenges was followed by *lack of knowledge* (i.e., education on preservation), and the *complexity of objects*. The complexity of objects includes challenges

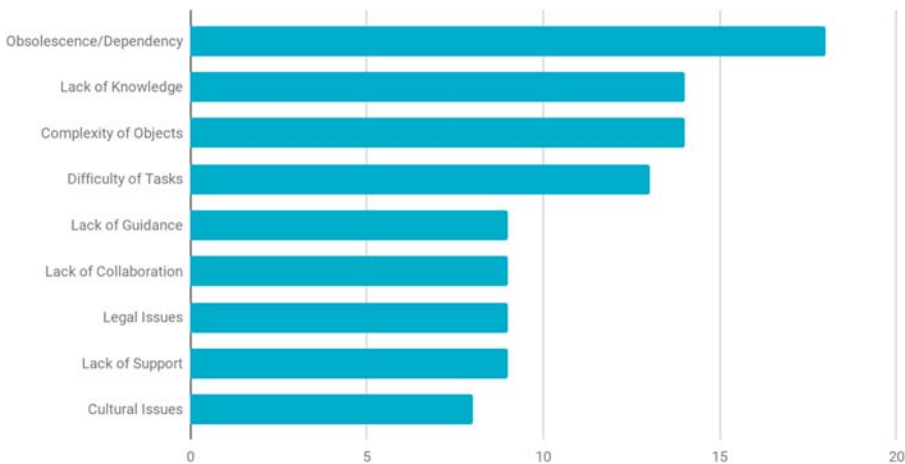


Figure 4. Challenges associated with sustaining access to dynamic data visualisations.

associated with objects that are composed of many layers (Broussard and Boss 2018) may be dependent on external resources such as APIs (Boss and Broussard 2017) or embedded resources (Brunelle et al. 2015, 2016, 2017). Several studies noted that knowledge about which objects to preserve in the first place was lacking (Barwick, Dearnley, and Muir 2011; Broussard 2015a), or understanding how to deal with versions (i.e., which version of an object is definitive? how many versions to preserve?) Choosing versions can be complicated by the dynamic nature of objects and by practices like personalisation or A/B testing (Broussard and Boss 2018). The types of digital objects discussed in these studies, including news apps and digital games, are complex and dynamic, and this complexity poses a major challenge to their preservation.

The *difficulty of performing preservation tasks*, in terms of the time, effort, or skills required (McCain 2015) was followed by *lack of guidance*, which is distinguished from *lack of knowledge* based on a stated absence of institutional guidelines or policy documents that outline preservation procedures or best practices. The noted absence of guidance suggests an awareness of the need for preservation actions coupled with the lack of organisational or professional frameworks for carrying it out, and was referred to in a third of the articles.

Closely related is *Lack of support*, which includes lack of funding, policy support, or human resourcing for preservation. This topic arose particularly frequently in the articles which dealt specifically with news preservation, occurring under various guises (Broussard 2015a; Broussard and Boss 2018; McCain 2015; Moore and Bonnet 2015).

Lack of collaboration was discussed in terms of a lack of knowledge sharing, assistance, or communication between parties involved in preservation. Clarity around who is responsible for the preservation, when many collaborators are involved in the production and publication of a data journalism output, also surfaced in a number of the articles. In the context of news preservation, Moore and Bonnet's survey found that, when asked who should be responsible for preserving born-digital news, "news workers and researchers more often selected libraries rather than archives; the opposite was true for library and archival staff" (Moore and Bonnet 2015, 356). In other fields, Barwick, Dearnley, and Muir (2011) observed that computer games preservation is often left to the gaming community, as there is uncertainty over what institutions should be responsible for preserving this content.

Legal issues were a challenge noted in over a quarter of studies, where preservation actions may infringe on copyright and intellectual property rights. This is particularly true of digital game preservation as the producers and copyright holders of digital games are often not the ones preserving their products (Barwick, Dearnley, and Muir 2011; Monnens et al. 2009; Takhteyev and DuPont 2013). Barateiro et al. (2012) also raised the issue of the prevalence of Software as a Service (SaaS) products and how copyright issues there may prevent preservation workflows.

Cultural issues around the value of preservation are also discussed as a challenge in the literature. For instance "the preservation of software is frequently seen as a secondary activity and one with limited usefulness" (Matthews et al. 2010, 92). Obstacles in this category may be related to a lack of education or advocacy around digital preservation within these communities.

Methods of Preservation

In addition to understanding the challenges and/or barriers to preserving and providing sustained access to dynamic data visualisations, the literature addresses commonly recommended or implemented preservation methods (see [Figure 5](#)).

Emulation is by far the most commonly discussed method of preservation, including the use of virtual machines, with *migration* coming second as a method.

Next is *documentation*, which involves collecting or preserving contextual materials outside of the digital object itself, such as information around the creation of the object, manuals, data associated with the object and related ephemera. This is followed by the *creation of snapshots*, which includes both web archiving methods (where a “snapshot in time” is taken of a website), and the capture of static images or flat versions of digital objects, like a “snapshot picture”, or a “screen grab”.

An important category emerged which covers the discussion of *specific tools* for preservation. The tools used for this purpose mentioned included *ReproZip* (Broussard and Boss 2018; Chirigati et al. 2016; Rampin et al. 2016; Steeves, Rampin, and Chirigati 2018), although all but one (Broussard and Boss 2018) of these papers were written by the team behind *ReproZip* for the purposes of introducing or demonstrating the tool. Other tools mentioned were *WebRecorder* (Broussard and Boss 2018), *RISA* (Remote Interface to Science Analysis), tools to decouple data processing software from infrastructure life-cycles (Gabriel et al. 2015; Ibarra and Gabriel 2012), media migration tools such as *DisFerret* (Takhteyev and DuPont 2013), and a variety of preservation, workflow management and configuration management tools (Chirigati et al. 2016; Steeves, Rampin, and Chirigati 2018).

Archiving and maintenance were the next two categories discussed in the literature. *Archiving* includes discussions around storage in archives and repositories for the purpose of long-term preservation. The category of *maintenance* covers the method of preserving an object “as-is” or trying to preserve an object by maintaining the hardware or software which supports the object over time.

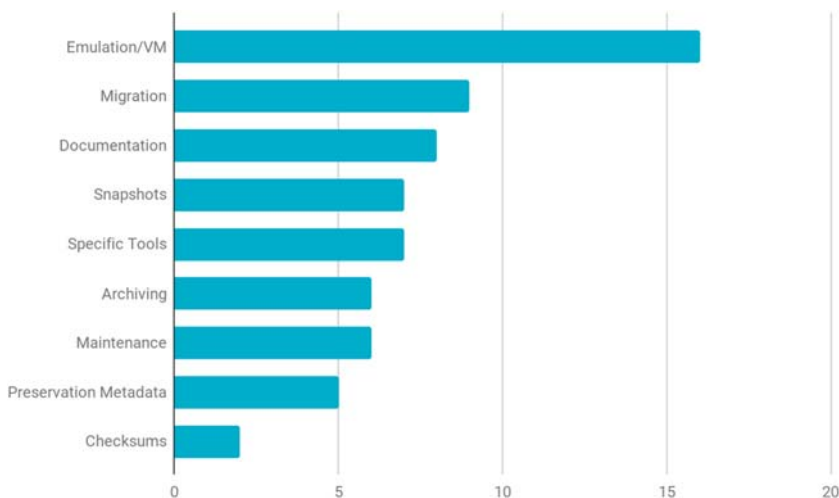


Figure 5. Methods of preservation.

The final methods noted are *Preservation Metadata* and *checksums*. Preservation metadata can be seen as related to both the documentation and archiving methods, and refers to storing information about the creation of the visualisation alongside the visualisation itself. Di Cosmo and Zacchiroli (2017) and Hildreth (2014) suggest the use of *checksums* for maintaining the integrity of the preserved object over time. Checksums are a form of computer-generated digital fingerprint of the bitstream of a file which can be used to detect whether that file has changed, for example, due to a copying error or to corruption of the storage media.

Recommendations for Preservation in the Literature

The recommendations around digital preservation were diverse in the literature, with thirteen different categories emerging. The frequency of these themes is depicted in Figure 6.

Tools occurred the most, with papers recommending specific tools, or advocating for tool development. The prevalence of tools is likely due to the fact that several of the articles included were from teams responsible for the development of these tools, as noted in the section above on Methods of Preservation, so the attention paid to them in the population may overrepresented. This does, however, suggest that there is significant activity surrounding the development of such tools.

The need for *guidance* in terms of best practice guidelines, common frameworks, preservation models, or further research in the area includes a call for “a report on the status of news apps” (Boss and Broussard 2017, 151), the stated goal of “promot [ing] a set of best practices for preservation ‘hygiene’” (Di Cosmo and Zacchiroli 2017, 4), and the idea of creating a “general model” for emulation of computer systems (Jamraj and Huang 2016, 4). In some cases, the articles aimed to provide frameworks or guidelines for preservation, recommending their use by others (Rios 2016) or stated their aim to develop such guidelines (Di Cosmo and Zacchiroli 2017; Meyerson et al. 2017).

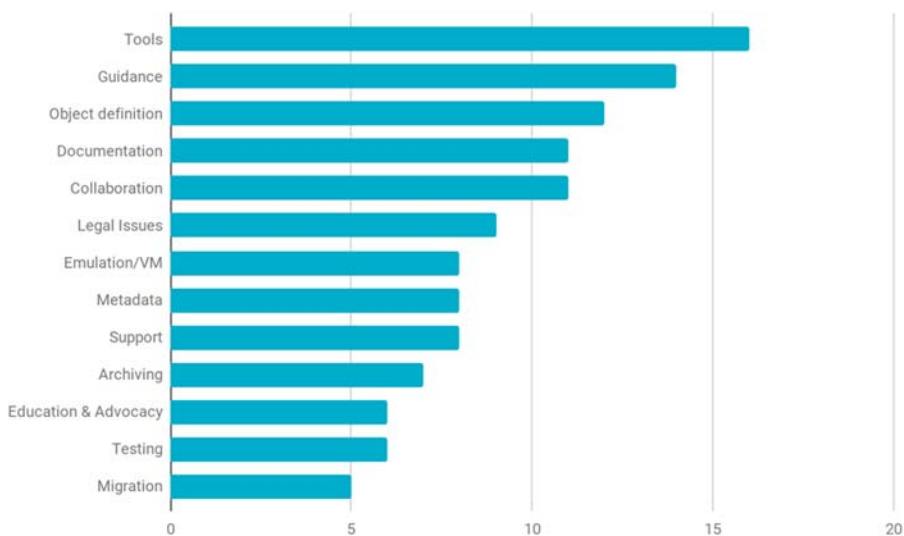


Figure 6. Recommended actions for preservation in the literature.

Object definition encompasses all recommendations related to the analysis of the preservation object. This includes issues of identifying objects to preserve, defining preservation objects, or deciding which versions of objects to preserve. Recommendations in this category included the need for works to be “archived at multiple stages ... so as to capture all the relevant versions possible” (Broussard and Boss 2018, 1210), the creation of an “online registry of news archives” in order to “assist in the assigning of resources to the most threatened news archives” (McCain 2015, 341), use of “the practice of software archaeology ... to reveal patterns in underlying architectures, preserve archetypal types of code-based art, select artifacts for preservation” (Marchese 2013, 25) and methods of collection development such as encouraging the donation of software or contextual information around apps (Monnens et al. 2009, 162; Takhteyev and DuPont 2013, 366).

Documentation recommends the preservation of contextual information about the object as well as the object itself, beyond basic metadata. Grad (2003) and Monnens et al. (2009) recommend the collection of oral histories in order to give context to preserved items. This has also been raised by Knight Lab Fellow, Fisher (2014), who suggests that “things like oral histories and screencasts will likely be required to tell future news developers and historians how this kind of journalism came to be and why we made the decisions we made”. Other recommendations include supplementing preserved material with static snapshots in the form of HTML or TIFF files (Boss and Broussard 2017, 152), preserving recordings and screen captures (Lowood 2011), and keeping physical artefacts, like original hardware (Rios 2016) or marketing ephemera and the writings and interviews of developers (Shustek 2006).

On par with documentation, the topic of *collaboration* arises in nearly a third of the reviewed articles. This includes recommendations to improve communication and increase collaboration between the academic community, industry and heritage institutions (Barwick, Dearnley, and Muir 2011), and maintaining a relationship with legal advisors and “the software publishing community” (Meyerson et al. 2017). Additionally, the studies on the preservation of digital games also emphasised the need to work with those who are already preserving software as a hobby or personal project, and to involve these people in preservation projects and provide them with education and resources where possible (Monnens et al. 2009; Takhteyev and DuPont 2013).

Several articles call for guidance on *legal issues*, so this theme was prevalent enough to warrant its own category. Meyerson et al. (2017), for instance, describes the ongoing collaboration with Harvard Law School’s Cyberlaw Clinic and their hope to “develop public-facing information sheets directed towards a practitioner audience, that provide overviews of legal topics relevant to software preservation”. In the area of news preservation, recommendations include “succession agreements” (McCain 2015, 346), “storing copyright and/or licensing information alongside news apps” (Broussard 2015a, 308) and “further work on establishing a legal pathway to the preservation of news apps” (Boss and Broussard 2017, 153). Barwick, Dearnley, and Muir (2011) and Broussard and Boss (2018) also recommend that legal deposit of digital objects be considered, analogous to the requirement to deposit printed material in a designated legal deposit library, which is typically mandated by law in most countries. This could provide a means of collection gathering for preservation, and circumvent the need to negotiate copyright issues on a collection-by-collection basis. The use of clearinghouses for storage of material which cannot be made publicly available is also recommended (Monnens et al. 2009).

Emulation and *migration* are two other categories that emerged in a considerable number of studies as recommendations for long-term preservation and access to such complex digital objects. Emulation appears slightly more popular, and Broussard and Boss (2018) suggest that this may be due to the greater effort involved in migration:

The effort needed to simply maintain dynamic digital objects online, and the amount of time, work, and organization involved in migrating or upgrading the websites themselves and the software environment required for them to display and function is impractical, if not unfeasible. Software packages and computational platforms change vastly over time. For these reasons, digital archivists believe that to save dynamic content such as data journalism for the long term, we must emulate, not migrate, the object in its computational environment. (Broussard and Boss 2018, 1209)

Studies which recommend *migration* cover software preservation (Rios 2016), game preservation (Monnens et al. 2009), general software preservation (Meyerson et al. 2017), the preservation of automatic test equipment (Dunn 2005) and of software relating to business processes (Barateiro et al. 2012). However, none of these studies specifically discuss migration as a recommendation for the preservation of news media-related objects.

Along with recommendations for both migration and emulation, Rios (2016) provides recommendations for *archiving software "as-is"* – in its current form – and maintenance, depending on the situation and type of object being preserved.

Recommendations relating to *metadata* call broadly for the use of metadata in preservation (Broussard 2015a; Marchese 2013; Matthews et al. 2010; Meyerson et al. 2017) or for more work on developing or choosing metadata structures (Hildreth 2014; Rios 2016). Di Cosmo and Zacchiroli (2017) described the types of information that should be included in the metadata, such as the programming language, version information, and licensing information for software artefacts. Boss and Broussard provide the most detailed recommendation for the adoption of a metadata schema, suggesting that PREMIS (Preservation Metadata Implementation Strategies) is "particularly applicable to news app archiving" (2018, 154).

Articles which make recommendations in the category of archiving included suggestions that archives or repositories be used or established for long-term preservation (Barateiro et al. 2012; Di Cosmo 2018; Di Cosmo and Zacchiroli 2017; Matthews et al. 2010). Broussard (2015a) recommends that news archiving should be built into existing systems and workflows so that items are archived automatically. The idea of centralised storage or storage in remote locations was also raised Dunn (2005), as was the use of "dark archives" for the preservation of objects which cannot be made publicly available (Monnens et al. 2009).

The category of *support* includes all recommendations related to increased funding, human resourcing, or policy support. Moore and Bonnet (2015) recommend that time, financial support and qualified personnel are needed to support news preservation. McCain (2015) notes that changes to the political environment to produce policy and funding changes will be necessary, along with the creation of financial incentives for news preservation, and support for new business models and grant writing. Others also call for further work on developing and investigating policy support (Hildreth 2014; Meyerson et al. 2017; Monnens et al. 2009; Rios 2016).

Education and *advocacy* include the provision of “marketing and training opportunities” in news preservation (McCain 2015, 340), emphasis of “broader and deeper technical competencies by librarians and archivists that enable them to engage with creators of complex digital objects” (Meyerson et al. 2017), and the provision of education in digital preservation for amateur game preservation communities (Takhteyev and DuPont 2013, 366). These recommendations encourage education for several different communities, including news workers, staff of memory institutions, and preservation enthusiasts, as well as promoting marketing and advocacy work.

Another category of recommendations that emerged is that of *testing*. This category includes recommendations related to the maintenance of authenticity and integrity of digital objects as they are being kept in long-term storage. It further includes recommendations to test web archives for missing embedded resources (Brunelle et al. 2015, 2016; Kelly, Nelson, and Weigle 2014) and the need for development of validation methods for software preservation (Barateiro et al. 2012).

Discussion

The findings in this study identify several challenges, ranging from specific technical challenges to broader social and organisational issues. Many preservation methods are described, however, there appears to be little guidance or knowledge about how to apply these. It is important to note suggested methods are not mutually exclusive, and a combination of methods might be selected as appropriate for any given visualisation.

Two broad approaches emerge from the preservation methods identified by this literature review. The first approach is to keep a working version of the visualisation available via methods such as emulation, migration, and virtual machines. The second approach attempts to capture a “flat” or simplified version of the visualisation via methods such as snapshots, documentation, and metadata. A flat or simplified version, considered in digital preservation language under the category of “surrogates”, essentially turns dynamic visualisations from complex digital objects into simple objects, which are more easily preserved. The dynamism is not maintained, but an effort is made to capture a sense of the original visualisation to preserve at least some part of it from total loss.

Selecting a Technical Preservation Approach

In choosing which of these approaches is most suitable for a given dynamic data visualisation in news and journalism, we draw on the concept of “Significant Properties”, proposed by Hedstrom and Lee (2002). Significant Properties are those properties of digital objects that affect their quality, usability, rendering, and behaviour. These are typically technical or behavioural characteristics of the digital objects, which need to remain unchanged when the file is accessed in the future, in order for the file to fulfil its original purpose. In the case of image files this might include aspects such as the height, width and colour depth of the image, while for video content it could include aspects such as the playback length and framerate. Software and other interactive digital objects tend to have more complicated significant properties relating to their behaviour and the types of user interaction possible.

Computer games, for example, are inherently experiential: the experience of the game is a significant property of the application. This can also be the case with data visualisations – that user interaction and experience are key to their meaning and value. Where this is the case, an approach using techniques such as emulation or migration may be indicated, as this preserves a working version of the original visualisation, and is thus more likely to preserve all of the significant properties of the object. This is in line with recent recommendations by the Digital Preservation Coalition on preserving Software (Morrissey 2020).

For some interactive data visualisations, dynamism and interactivity are not significant properties of the object: much is translated without these aspects. In such cases, it may not be necessary to preserve the entire interactive data visualisation in a working form, as an approach using snapshots and documentation as surrogates for the original may satisfactorily retain the significant properties.

Visualisations in data journalism are in most cases web-based, and have similar preservation issues as other web pages. We can thus look to web page significant properties for further guidance. Table 1 provides an extract of the significant properties identified by van Veenendaal, Lucker, and Sijtsma (2018) for interactive web pages.

We are particularly interested in the properties related to dynamically-created content, such as “JavaScript scripting for interactivity”, “Embedded interactive content encoded in other formats (e.g., Flash, SVG)” and “Dynamism (is the content static or generated on demand)”. Identifying whether and to what extent these are significant properties of a visualisation can help in selecting which approach to take in its preservation. Other theoretical approaches in the area of software preservation, such as the codifying and identification of levels of interactivity (Schulmeister 2003; Stalph and Heravi 2021), may also be instructive and should be targets of future research.

Although such theoretical frameworks may provide a starting point to selecting the appropriate preservation approach, they are only a partial guide. Dynamic data visualisations, although primarily web-based and sharing many significant properties with other web applications, are not identical to websites, in that they have unique aspects when considering preservation as a process. For example, there is a variance in the tools used by journalists within and between organisations, and multiple points of input to the organisation’s CMS. Using iFrames is one of the common ways to embed data visualisations created with external online tools into stories. An iFrame essentially creates an opening on an HTML page, which can pull content from external websites, including visualisations created in a range of external websites, such as Datawrapper and Flourish. Most

Table 1. Significant Properties for web pages: extract from van Veenendaal, Lucker, and Sijtsma (2018).

Information type	Category	Significant property
Web page	Appearance	Basic (X)HTML textual formatting for presentation
Web page	Appearance	Cascading Style Sheets (CSS) for presentation
Web page	Behaviour	(X)HTML (hyper)links
Web page	Behaviour	(X)HTML forms
Web page	Behaviour	JavaScript scripting for interactivity
Web page	Behaviour	Embedded interactive content encoded in other formats (e.g., Flash, SVG)
Web page	Behaviour	Dynamism (is the content static or generated on demand)
Web page	Content	Textual content

online data visualisation tools provide iFrame embed codes, which the journalist can simply copy and paste to their organisational CMS. The content remains hosted externally, but appears in the opening or portal on the news site's page, which means if the external content is moved or removed, it may disappear or show an error on the news site. [Figure 1](#) depicts a real-life example of such scenario.

Well-developed tools exist to capture entire webpages or websites (Pope and Beresford 2007), but they are not always able to capture highly interactive data visualisations, or those that rely on server-side applications and data. Furthermore, capturing the web through this method (which creates Web Archiving – WARC – files) is difficult and not likely to be implemented as part of journalistic workflows. Other tools have been developed with the specific aim of capturing or archiving a variety of news media such as Story Tracker (Welsh 2015), Save My News (Welsh 2017), Gifinator (BBC News Labs 2016) or RadioReader (BBC News Labs 2017). These use a variety of web-archiving and snapshotting approaches and are relatively easy to use. These are not, however, specifically aimed at data visualisations and they have mixed results when capturing interactive content.

While it may be most desirable to preserve all visualisations in their original working form, emulation and migration approaches are more complex and may require more frequently attention over time. There may also be issues gaining permission or access to emulate or migrate, in particular where the visualisation is externally hosted. The method of identifying significant properties is useful, but must be considered alongside larger resource and workflow requirements for preservation.

Dealing with Non-technical Challenges

Regardless of the technical approach taken to preservation, several systemic methods emerged from the literature review that draw on recognised topics in the digital preservation domain. These include archiving, the creation of preservation metadata and checksums, integrity testing and so on. These aspects must be considered in conjunction with any decisions on the overall approach as described above and involve saving and preserving either the original version or the simplified surrogate and ensuring ongoing integrity and access to that saved object.

Trustworthy Digital Repositories handle many of the actions required for preservation. A Trustworthy Digital Repository is one which has been certified (e.g., under CoreTrustSeal or ISO16363) as having the necessary policies, procedures and processes in place to ensure that deposited material will remain accessible for the long term.

Overall, the challenges identified in this study indicate that the complexity of the task of preservation is the biggest obstacle to preserving these objects. This complexity is not limited to technical aspects. Rather, it is in part attributable to the wider cultural or organisational challenge of digital preservation, where resources – financial and human – are limited, preservation is not embedded in publication workflows, and advocates for preservation are few and far between. Broussard (2015a, 309) notes that the “reason that news apps don’t get archived at legacy media organizations has to do with the back-end technology of the newsroom”, as archiving systems usually pull content from the print CMS rather than the web CMS. Speaking about his experience as a senior software architect at the New York Times, (Pennock 2013) states that “almost any news programmer generally loathes their organisation’s Content Management System ... And so, we do our work

outside the CMS, skinning our pages to look like the main news site while remaining architecturally apart". If proper preservation is to be undertaken by the creators of news apps or staff at news organisations, the realities of current workflows and productivity blockers must be kept in mind, with preservation actions integrated into existing workflows as seamlessly as possible. Furthermore, the responsibility for these actions must be identified and pursued systematically (Delve, Denard, and Kilbride 2014).

Awareness-building around preservation, guidelines for preserving visualisations, and training on how to integrate preservation into workflows can assist with these larger social or organisational challenges.

Recommendations

The following recommendations draw on the systematic study of the literature, methods detailed in professional literature from the fields of data journalism and digital preservation, as well as our professional expertise as academics and practitioners in these areas. There is a significant body of professional guidelines addressing different areas of digital preservation that suggests immediate steps. A few examples include the Digital Preservation Handbook (DPC 2015), the "Preserving Digital Collections" pages on the website of the National Archives, UK (National Archives, UK n.d.) and pages on Digital Preservation at the Library of Congress (Library of Congress n.d.).

Our recommendations below fall into several categories, and many of these require changes to workflows and investment into new policies, practices and technical solutions, and will therefore take time to pursue. To mitigate potential loss in the more immediate term, following these recommendations we also provide concrete actions that do not require major workflow changes.

Guidance and Education

Many of the publications studied here call for guidelines, or state their intention to develop them, but only a few provide concrete frameworks or guidelines for preservation (Jamraj and Huang 2016; Matthews et al. 2010; Rios 2016). There is a clear need for guidance across a range of preservation-related topics. These include:

- Guidance on selecting an appropriate approach for preservation based on established Digital Preservation theory such as Significant Properties, for example, Significant Properties lists, decision trees, etc.
- Tools and guides to simplify the technical process of emulation or migration of highly interactive dynamic visualisations
- Recommendations on the tools and processes for creating snapshots, the recommended formats for snapshots (videos, screencaptures, animated GIFs, web archiving, etc.) and the types of documentation (written documentation, oral histories, etc.). A number of screen capture tools exist, such as the Microsoft Windows Game Bar application or Apple QuickTime. Some other free screen capture looks such as Apowersoft Online Screen Recorder³ or ShareX⁴ can automatically save the screen recording as an animated GIF file.

- Advice around the legal and permissions issues relating to the export and archiving of visualisations, in particular those hosted externally to the news organisation. (Particularly relevant where an emulation or migration approach is required).
- Assistance in identifying suitable Trustworthy Digital Repositories for long-term preservation of dynamic visualisations, or advice on building these within the news organisation.
- Community-based guidance around making data visualisations FAIR (Wilkinson et al. 2016), so that they are not merely deposited, but are made sustainably available through the inclusions of persistent identifiers, sufficient metadata and contextual information, and so on.

Finally, digital preservation topics are to a great degree absent from data journalism courses (Heravi 2017, 2019) and should be added urgently.

Infrastructures and Tools

There is a need for Trustworthy Repositories for digital news media, including dynamic visualisations. Larger national media outlets typically have in-house archiving systems which are used to preserve their textual, image and audiovisual digital content. There is an urgent need to extend these platforms to provide archiving for dynamic data visualisations. Alternatively, such content could be deposited in shared domain repositories, or the repositories of national memory institutions with a mandate to preserve the public record.

The creation of emulated or migrated versions of a dynamic data visualisation may currently be hampered by a lack of export functionality in the popular tools used to create and host visualisations. Vendors should be encouraged to provide export functionality, journalists should aim to choose these vendors and tools. We have discussed this further in the next sub-section, as part of specific recommendations for journalists. In addition, the preservation community can help by identifying or creating tools that support the creation of emulated and migrated versions of data visualisations from vendor export files. Open-source tools can be used to avoid vendor lock-in.

Collaboration

Digital preservation is a process, and not simply an action that is taken once and then forgotten. It requires active, ongoing maintenance. In the long term, news organisations need to fully integrate preservation into production workflows, so that preservation is not an after-thought. Advocating for a preservation policy and dedicated resources for preservation is one way forward, but in the nearer term, news organisations can also partner with libraries, archives, or preservation infrastructures to build internal capacity, provide proof of concept, and develop some of the steps that will integrate preservation at an organisational level. This effort will benefit the archival work already undertaken by most news organisations and provide a better understanding of how the growing aspect of dynamic data visualisations can be preserved alongside more traditional outputs. Organisations can look to professional bodies, such as the Digital Preservation Coalition (DPC 2015) for guidance.

Where collaboration with memory institutions is pursued it will be necessary to clearly identify who is responsible for this preservation activity and in which organisation it should take place. Collaborations with Trustworthy Digital Repositories, and with the Certification bodies for such repositories is also strongly recommended as such bodies can provide concrete assistance as well as advice on best practice for preservation, even where the news organisation is not pursuing certification for their own repository.

Funding and Resourcing

Funding and resourcing are required to enable the systematic creation and maintenance of emulated and migrated versions of data visualisations, as well as for archiving and preserving these in the long term. This will not come about without consistently highlighting the dangers of the loss of dynamic data visualisations and advocacy for their preservation.

Legal Frameworks

If data visualisations are to be deposited in external repositories or the infrastructures of memory institutions, clarity around copyright and licencing conditions must be sought. Ideally, legal frameworks for deposit, analogous to current Legal Deposit requirements for printed material should be explored. This task may involve changes to legislation as well as a collaborative dialogue between current Legal Deposit Libraries, news media organisations, copyright experts within relevant jurisdictions, and the broader digital preservation community, and would obviate many of the complications relating to permissions, rights clearance and other legal issues. In the short term, guidance on rights clearance and access issues should be developed with the intention of preservation and continued access in mind. In the longer term, this complex topic merits considerable future research, and should be a focus of future collaborations between news media organisations and memory institutions.

The above recommendations address the need for an organised and sustainable approach to the long-term digital preservation of data visualisations. They aim to ensure that these increasingly important elements of journalistic output are routinely preserved alongside simpler forms of digital news media. It is important to note, however, that these are all medium- to long-term actions which require investment of effort over time, financial resources, and collaborations that may expand the remit of existing institutions.

Recommendations for Immediate and Practical Interventions by Data Journalists

In the more immediate future, there are also simple actions data journalists can take to ensure partial preservation of their content in lieu of more robust approaches. These recommendations are based on the literature review, a study of key professional and/or non-peer-reviewed contributions, as well as the experience of the authors working as, and with, data journalists. These are further inline with recommendations by Broussard (2020).

An approach that assumes limited time and resources combines a basic identification of significant properties, along with the creation of surrogate output of types that are

easily preservable using current technologies, such as images and audiovisual formats. This is essentially a basic form of the *snapshot* method identified in the literature.

We propose that for every dynamic data visualisation included in a story, the journalist should:

- (1) Identify the significant properties of the data visualisation, in terms of the importance for the story at hand.
- (2) If an image screenshot of the data visualisation could represent these properties to a satisfactory degree, then take a screenshot of the visualisation, and store it with other archived audiovisual content.

Screenshots have been used by some news organisations in their archiving practices. [Figure 7](#) depict two examples from Washington Post and the New York Times, where the story is missing due to the issue of Flash, but the organisations offer access to alternate archived content.

Following the link in the Washington Post story in [Figure 7](#) retrieves a PDF, which had been previously generated for the print version of the story. Clearly, this conveys an acceptable degree of the original story's intention. However, the link in the NYT story retrieves a screenshot that only shows the first slide of a multi-slide story, which means a significant part is missing.

- (1) If an image screen grab cannot capture the story to a satisfactory level, then we propose two alternatives:
 - (a) If a small number of screen grabs can tell the story, then create a GIF animation that includes these in sequence, and archive as above. GIF animations allow limited animation but are nonetheless relatively simple image files which are straightforward to preserve. Many news organisations already create animated GIFs for content promotion on social media and so the tools and expertise are readily available, for example, the Economist Data desk (Segger 2018).
 - (b) If a number of screen grabs in the form of a GIF animation cannot do justice to the visualisation, then consider creating a video cast of the data visualisation in use, highlighting the most important parts. A range of widely available free tools can be used to create such video content which is also relatively simple to preserve.

These simple surrogate representations must also be linked to the original story to ensure that the reader can find them if the story remains available, but the original visualisation is no longer available. This linking could be via a structural solution whereby the CMS of the news organisation allows an alternate link to be specified, and automatically displays the file behind the link if the main visualisation fails to load. An alternative or possible interim solution would be to include a link under each visualisation to the surrogate version which invites the user to click on it if the visualisation does not display correctly. An example of how this has worked in practice could be seen in [Figure 7](#).

Creating an image, GIF animation or a video of your data visualisation is an uncomplicated solution that enables the capture of significant properties in terms of content as story, providing a stop-gap until more systematic and sophisticated methods for

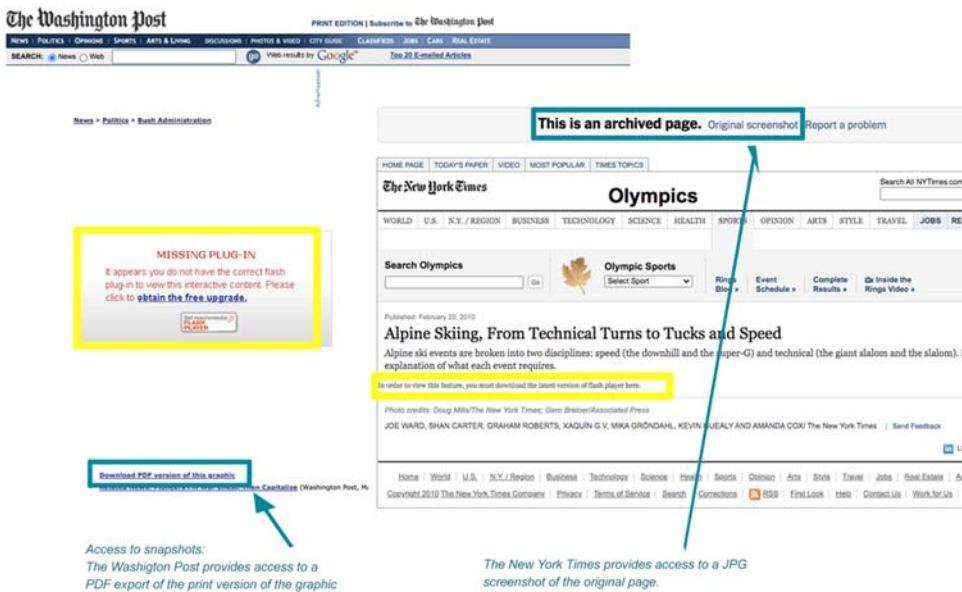


Figure 7. Screenshots from Washington Post and New York Times, depicting the disappearance of the stories due to the deprecation of Flash, as well as the accessible screenshots through their archives: www.washingtonpost.com/wp-srv/politics/pioneers/pioneers_spheres.html and archive.nytimes.com/www.nytimes.com/interactive/2010/02/20/sports/olympics/downhill-overview.html. Screenshots taken on 21st February 2021.

preservation of dynamic data visualisations are in place. In addition to long-term preservation and access, this simple method could also cater for issues associated with loading complex objects across devices.

As such, we also propose that every provider of data visualisation creation tools should ideally provide GIF animation and video exports, in addition to their current visualisation exports. Many data visualisation providers promise their users that in the case of company closure, users will be given the option to download the code behind the charts. This is a responsible offer, but most journalists will not have the time or skills to execute that code on a different platform. Nor will they be able to go back to every single story they created to update the server information for where the data visualisation is hosted. Hence, it is advisable that journalists create simple exports of their data visualisations at time of publication, and provide the information for how these can be accessed if the original publication fails. Both data journalists, and the wider digital preservation community, should advocate with vendors of these tools to help bring this about.

These immediate and relatively contained measures could ensure that much of the data journalism currently being produced is not lost entirely, while the newsrooms find ways to implement the recommendations to ensure longer-term systematic preservation of such complex objects.

As a final note, we would like to emphasise that digital preservation is an ongoing process, not simply an endpoint, and that methods must evolve within and by the communities that are most invested in the long-term stewardship of their outputs. Because of journalism's fundamental and unique contribution to the historical record, it is imperative

that preservation is built into the production of data journalism, so that this key element of the record is not lost.

Notes

1. Datawrapper and Flourish are two online data visualisation tools popular amongst data journalists. Several others exist, such as Infogram, Google Charts, Tableau and CARTO. A curated list of tools commonly used by data journalists can be found at <http://bit.ly/BHddjtools>.
2. See the following Interest or Working Groups: RDA/WDS Certification of Digital Repositories <https://www.rd-alliance.org/groups/rdawds-certification-digital-repositories-ig.html> which produced the TRUST Principles (Lin, D., Crabtree, J., Dillo, I. et al., 2020), the Preservation Infrastructure IG <https://www.rd-alliance.org/groups/preservation-e-infrastructure-ig.html>, the Preservation Tools, Techniques, and Policies (PTTP) IG <https://www.rd-alliance.org/groups/preservation-tools-techniques-and-policies>. The Software Source Code IG <https://www.rd-alliance.org/groups/software-source-code-ig>
3. <https://www.apowersoft.com/free-online-screen-recorder>
4. <https://getsharex.com/>

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Bahareh Heravi  <http://orcid.org/0000-0002-0587-1046>

Kathryn Cassidy  <http://orcid.org/0000-0002-0646-0628>

Edie Davis  <http://orcid.org/0000-0003-4043-0358>

Natalie Harrower  <http://orcid.org/0000-0002-7487-4881>

References

- Barateiro, J., D. Draws, M. A. Neumann, and S. Strodl. 2012. "Digital Preservation Challenges on Software Life Cycle." In *2012 16TH European Conference on Software Maintenance and Reengineering (CSMR)*, edited by T. Mens, A. Cleve, and R. Ferenc, 487–490. IEEE. <https://doi.org/10.1109/CSMR.2012.63>.
- Barwick, J., J. Dearnley, and A. Muir. 2011. "Playing Games with Cultural Heritage: A Comparative Case Study Analysis of the Current Status of Digital Game Preservation." *Games and Culture* 6 (4): 373–390. <https://doi.org/10.1177/1555412010391092>.
- BBC News Labs. 2016. *Gifinator*. <https://bbcnewslabs.co.uk/projects/gifinator/>.
- BBC News Labs. 2017. *RadioReader*. <https://bbcnewslabs.co.uk/projects/radioReader/>.
- Berman, F. 2008. "Got Data? A Guide to Data Preservation in the Information age." *Communications of the ACM* 51 (12): 50–56. <https://doi.org/10.1145/1409360.1409376>.
- Boss, K., and M. Broussard. 2017. "Challenges of Archiving and Preserving Born-digital News Applications." *IFLA Journal-International Federation of Library Associations* 43 (2): 150–157. <https://doi.org/10.1177/0340035216686355>.
- Broussard, M. 2015a. "Preserving News Apps Present Huge Challenges." *Newspaper Research Journal* 36 (3): 299–313. <https://doi.org/10.1177/0739532915600742>.
- Broussard, M. 2015b. The Irony of Writing Online About Digital Preservation. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2015/11/the-irony-of-writing-about-digital-preservation/416184/>.

- Broussard, M. (2020). Archiving Data Journalism. In *The Data Journalism Handbook: Towards a Critical Data Practice*. European Journalism Centre. <https://datajournalism.com/read/handbook/two/organising-data-journalism/archiving-data-journalism>
- Broussard, M., and K. Boss. 2018. "Saving Data Journalism: New Strategies for Archiving Interactive, Born-digital News." *Digital Journalism* 6 (9): 1206–1221. <https://doi.org/10.1080/21670811.2018.1505437>.
- Brunelle, Justin F, M. Kelly, H. Salaheldeen, M. C. Weigle, and M. L. Nelson. 2015. "Not all Mementos Are Created Equal: Measuring the Impact of Missing Resources." *International Journal on Digital Libraries* 16 (3–4): 283–301. <https://doi.org/10.1007/s00799-015-0150-6>.
- Brunelle, Justin F, M. Kelly, M. C. Weigle, and M. L. Nelson. 2016. "The Impact of JavaScript on Archivability." *International Journal on Digital Libraries* 17 (2): 95–117. <https://doi.org/10.1007/s00799-015-0140-8>.
- Brunelle, J F, Weigle, M. C., & Nelson, M. L. (2017). Archival Crawlers and JavaScript: Discover More Stuff but Crawl More Slowly. *Proceedings of the ACM/IEEE Joint Conference on digital libraries*. <https://doi.org/10.1109/JCDL.2017.7991554>
- Chirigati, F., R. Rampin, D. Shasha, and J. Freire. 2016. ReproZip: Computational Reproducibility With Ease. *Proceedings of the 2016 International Conference on Management of Data*, 2085–2088. <https://doi.org/10.1145/2882903.2899401>.
- Coddington, M. 2015. "Clarifying Journalism's Quantitative Turn." *Digital Journalism* 3 (3): 331–348. <https://doi.org/10.1080/21670811.2014.976400>.
- Delve, J., Denard, H., & Kilbride, W. (2014). Digital Preservation Strategies for Visualizations and Simulations. In *Preserving Complex Digital Objects*, edited by Janet Delve and David Anderson, 143–153. London, UK: Facet Publishing.
- Di Cosmo, R. 2018. Software Heritage: Why and how we collect, preserve and share all the software source code. *Proceedings of the 40th IEEE/ACM International Conference on Software Engineering: Software Engineering in Society*, 2.
- Di Cosmo, R., and S. Zacchiroli. 2017, September. *Software Heritage: Why and How to Preserve Software Source Code*. PRES 2017 – 14th International Conference on Digital Preservation, Kyoto, Japan. <https://hal.archives-ouvertes.fr/hal-01590958>.
- DPC. 2015. *Digital Preservation Handbook, 2nd Edition*. Digital Preservation Coalition. <https://www.dpconline.org/handbook>.
- Dunn, D. 2005. "Software Preservation via Media Migration." *EEE Autotestcon 2005*: 571–573. <https://doi.org/10.1109/AUTEST.2005.1609199>.
- The European Commission Expert Group on FAIR Data. 2018. *Turning FAIR Into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data*. Luxembourg: Directorate-General for Research and Innovation (European Commission). <https://doi.org/10.2777/1524>.
- Fink, A. 2005. *Conducting Research Literature Reviews: From the Internet to Paper (Second Edition)*. London, CA: Sage.
- Fisher, T. 2014, March 18. *Preserving interactive news projects with Newseum, OpenNews and Pop Up Archive*. Knight Ab. <https://knightlab.northwestern.edu/2014/03/18/preserving-interactive-news-projects-with-newseum-opennews-and-pop-up-archive/>.
- Gabriel, C., C. Arviset, A. Ibarra, and A. Pollock. 2015. "Long-term Preservation of Data Analysis Capabilities." In *Astronomical Data Analysis Software and Systems XXIV*, edited by A. R. Taylor and E. Rosolowsky, Vol. 495, 245–248. Astronomical Society of the Pacific.
- Grad, B. 2003. "Preserving the Software Industry's Past." *IEEE Annals of the History of Computing* 25 (1): 88–83. <https://doi.org/10.1109/MAHC.2003.1179890>.
- Hansen, K. A., and N. Paul. 2017. *Future-proofing the News: Preserving the First Draft of History*. Lanham, MD: Rowman & Littlefield.
- Harris, J. 2013. And Remember, this Is for Posterity. *OpenNews*. <https://source.opennews.org/articles/and-remember-ones-posterity/>.
- Hedstrom, M., and C. A. Lee. 2002. "Significant Properties of Digital Objects: Definitions, Applications, Implications." *Proceedings of the DLM-Forum 2002 @Ccess and Preservation of Electronic Information: Best Practices and Solutions*, 218–227.

- Heravi, B. 2017. "Teaching Data Journalism." In *Data Journalism: Past, Present and Future*, edited by J. Mair, R. Keeble, M. Lucero, and M. Moore, 221–228. United Kingdom: Abramis Academic Publishing.
- Heravi, B. 2019. "3WS of Data Journalism Education: What, Where and Who?" *Journalism Practice* 13 (3): 349–366.
- Heravi, B., and M. Lorenz. 2020. "Data Journalism Practices Globally: Skills, Education, Opportunities, and Values." *Journalism and Media* 1 (1): 26–40. <https://doi.org/10.3390/journalmedia1010003>.
- Hildreth, M. D. 2014. "Data and Software Preservation for Open Science (DASPOS)." *IEEE 30th International Conference on Data Engineering*, 142–146. <https://doi.org/10.1109/ICDEW.2014.6818318>.
- Ibarra, A., and C. Gabriel. 2012. "Solving the Software Legacy Problem with RISA." In *Astronomical Data Analysis Software and Systems XXI*, edited by P. Ballester, D. Egret, and N. P. F. Lorente, Vol. 461, 643–646. Astronomical Society of the Pacific.
- Jamraj, D., & Huang, S. (2016). Developing an emulation model towards the preservation of modern computing systems. 2015 *IEEE 2nd International Conference on Information Science and Security, ICISSE 2015*. <https://doi.org/10.1109/ICISSE.2015.7370996>
- Kelly, M., M. L. Nelson, and M. C. Weigle. 2014. "The Archival Acid Test: Evaluating Archive Performance on Advanced HTML and JavaScript." *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 25–28. <https://doi.org/10.1109/JCDL.2014.6970146>.
- Kitchenham, B., and S. Charters. 2007. *Guidelines for Performing Systematic Literature Reviews in Software Engineering (No. EBSE-2007-01) [Technical Report]*. Keele University and University of Durham. https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf.
- Library of Congress. n.d. *Digital Preservation at the Library of Congress*. Retrieved 12 January 2021, from <https://www.loc.gov/preservation/digital/>.
- Lin, D., J. Crabtree, I. Dillo, R. R. Downs, R. Edmunds, D. Giaretta, M. De Giusti, et al. 2020. "The TRUST Principles for Digital Repositories." *Scientific Data* 7 (1): 144. <https://doi.org/10.1038/s41597-020-0486-7>.
- Lowood, H. 2011. "Perfect Capture: Three Takes on Replay, Machinima and the History of Virtual Worlds." *Journal of Visual Culture* 10 (1): 113–124. <https://doi.org/10.1177/1470412910391578>.
- Lynch, C. 2008. "How Do Your Data Grow?" *Nature* 455 (7209): 28–29. <https://doi.org/10.1038/455028a>.
- Marchese, F. T. 2013. "Software Archaeology and the Preservation of Code-based Digital Art." In *Proceedings of the IS&T Archiving Conference 2013*, 25–30. Washington, D.C., USA.
- Matthews, B., A. Shaon, J. Bicarregui, and C. Jones. 2010. "A Framework for Software Preservation." *International Journal of Digital Curation* 5 (1): 91–105. <https://doi.org/10.2218/ijdc.v5i1.145>.
- McCain, E. 2015. "Plans to Save Born-digital News Content Examined." *Newspaper Research Journal* 36 (3): 337–347. <https://doi.org/10.1177/0739532915600747>.
- Melendez, S. 2018, December 11. "RIP Fusion Tables: Google is Killing off the Beloved Data Visualization Tool." *Fast Company*. <https://www.fastcompany.com/90279779/rip-fusion-tables-google-is-killing-off-the-beloved-data-visualization-tool>.
- Meyerson, J., Z. Vowell, W. Hagenmaier, A. Leventhal, F. Rios, E. R. Roke, and T. Walsh. 2017. "The Software Preservation Network (SPN): A Community Effort to Ensure Long Term Access to Digital Cultural Heritage." *D-Lib Magazine* 23: 5–6. <https://doi.org/10.1045/may2017-meyerson>.
- Monnens, D., Z. Vowell, J. E. Ruggill, K. S. McAllister, and A. Armstrong. 2009. "Before It's Too Late: A Digital Game Preservation White Paper." *American Journal of Play* 2 (2): 139–166.
- Moore, J. E., and J. L. Bonnet. 2015. "Survey Finds Differences on Preserving Born-digital News." *Newspaper Research Journal* 36 (3): 348–362. <https://doi.org/10.1177/0739532915600748>.
- Morrissey, S. M. 2020. *Preserving Software: Motivations, Challenges and Approaches*. Digital Preservation Coalition. <https://doi.org/10.7207/twgn20-02>.
- National Archives, UK. n.d. *Preserving Digital Collections*. Accessed January 12, 2021, from <https://www.nationalarchives.gov.uk/archives-sector/advice-and-guidance/managing-your-collection/preserving-digital-collections/>.

- OECD. 2007. *OECD Principles and Guidelines for Access to Research Data from Public Funding*. <https://www.oecd-ilibrary.org/content/publication/9789264034020-en-fr>.
- OECD. 2017. *Business Models for Sustainable Research Data Repositories*. OECD Science, Technology and Industry Policy Papers. No. 47. Paris: OECD Publishing. <https://doi.org/10.1787/302b12bb-en>.
- Ojo, A., and B. Heravi. 2018. "Patterns in Award Winning Data Storytelling: Story Types, Enabling Tools and Competences." *Digital Journalism* 6 (6): 693–718. <https://doi.org/10.1080/21670811.2017.1403291>.
- Okoli, C. 2015. "A Guide to Conducting a Standalone Systematic Literature Review." *Communications of the Association for Information Systems* 37), <https://doi.org/10.17705/1CAIS.03743>.
- Pennock, M. 2013. *Web-Archiving*. Digital Preservation Coalition. <https://doi.org/10.7207/twr13-01>.
- Pilat, D., and Y. Fukasaku. 2007. "OECD Principles and Guidelines for Access to Research Data from Public Funding." *Data Science Journal* 6: OD4–OD11. <https://doi.org/10.2481/dsj.6.OD4>.
- Pope, J., and P. Beresford. 2007. IIPC Web Archiving Toolset Performance Testing at the British Library. Ariadne, 52. <http://www.ariadne.ac.uk/issue/52/pope-beresford/>.
- Rampin, R., F. Chirigati, D. Shasha, J. Freire, and V. Steeves. 2016. "ReproZip: The Reproducibility Packer." *The Journal of Open Source Software* 1 (8): 107. <https://doi.org/10.21105/joss.00107>.
- Rios, F. 2016. "The Pathways of Research Software Preservation: An Educational and Planning Resource for Service Development." *D-Lib Magazine* 22: 7–8. <https://doi.org/10.1045/july2016-rios>.
- Schouppe, M., and J.-C. Burgelman. 2018, October 9. "Relevance of the EOSC Initiative and FAIR Principles in the Realm of Open Science and Implementation Phases of the EOSC." XX International conference "data analytics and Management in data intensive domains" (DAMDID/RCDL'2018), Moscow, Russia.
- Schulmeister, R. 2003. "Taxonomy of Multimedia Component Interactivity. A Contribution to the Current Metadata Debate." *New Media in Education - Special Issue of Studies in Communication Sciences* 3 (1): 61–80.
- Segel, E., and J. Heer. 2010. "Narrative Visualization: Telling Stories with Data." *IEEE Transactions on Visualization and Computer Graphics* 16 (6): 1139–1148. doi.org/10.1109/TVCG.2010.179.
- Segger, M. 2018. Lessons for Showcasing Data Journalism on Social Media. *Medium*. <https://medium.com/severe-contest/lessons-for-showcasing-data-journalism-on-social-media-17e6ed03a868>.
- Shustek, L. 2006. "What Should We Collect to Preserve the History of Software?" *IEEE Annals of the History of Computing* 28 (4): 112–111. <https://doi.org/10.1109/MAHC.2006.78>.
- The Silk Team. 2017. "It's Time to Say Goodbye." *Silk Blog*. <http://blog.silk.co/post/167155630197/its-time-to-say-goodbye>.
- Stalsh, F., and B. Heravi. 2021. "Exploring Data Visualisations: An Analytical Framework Based on Dimensional Components of Data Artefacts in Journalism." *Digital Journalism, To Appear*.
- Steeves, V., R. Rampin, and F. Chirigati. 2018. "Using ReproZip for Reproducibility and Library Services." *IASSIST Quarterly* 42 (1): 14. <https://doi.org/10.29173/iq18>.
- Takhteyev, Y., and Q. DuPont. 2013. "Retrocomputing as Preservation and Remix." *Library Hi Tech* 31 (2): 355–370. <https://doi.org/10.1108/07378831311329103>.
- UNESCO. 2020. UNESCO Builds an International Policy Agenda for Digital Preservation of Documentary Heritage. <https://en.unesco.org/news/unesco-builds-international-policy-agenda-digital-preservation-documentary-heritage>.
- van Veenendaal, R., P. C. M. Lucker, and C. D. Sijtsma. 2018. "Significant Significant Properties." <https://openpreservation.org/wp-content/uploads/2018/10/Significant-Significant-Properties.pdf>.
- Welsh, B. 2015. "Story Tracker." <https://storytracker.readthedocs.io/>.
- Welsh, B. 2017. "Save My News." <https://www.savemy.news/>.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3: 1. doi:10.1038/sdata.2016.18.